International Journal of Innovative Technology and Interdisciplinary Sciences
https://journals.tultech.eu/index.php/ijitis
ISSN: 2613-7305
Volume 8, Issue 3
DOI: https://doi.org/10.15157/IJITIS.2025.8.3.784-817
Received: 30.09.2025; Revised: 20.10.2025; Accepted: 24.10.2025

Research Article

Cluster Analysis of Merger and Acquisition Patterns in the Electronic Design Automation Industry Using Machine Learning Techniques

Blerim Zylfiu[1*], Galia Marinova[2*], Edmond Hajrizi[2], Besnik Qehaja[2]

[1] Computer Science and Engineering, University for Business and Technology, Pristina, Kosovo
[2] Faculty of Telecommunications, Technical University of Sofia, Sofia, Bulgaria
***blerim.zylfiu@ubt-uni.net, gim@tu-sofia.bg**

Abstract

This study develops a three-stage Reverse Hybrid Clustering system which detects Electronic Design Automation industry merger and acquisition patterns. The method unites DBSCAN density-based spatial clustering, K-means boundary refinement and noise reintegration to study 661 M&A deals from 1975- 2025. The Reverse Hybrid method produces a Silhouette Score of 0.8786 and Davies-Bouldin Index of 0.1144, representing a 201% improvement from K-means (0.2918) and outperforms Gaussian Mixture Models (0.162), Forward Hybrid (-0.0273), and DBSCAN (-0.107). Five M&A archetypes emerge across time periods, with target company operational development explaining 76% of variance (χ^2 = 534.36, p < 0.001). A composite metric, TechImpactScore, is introduced, combining patent portfolios (35%), deal values (30%), company maturity (20%), and acquirer activity (15%). This metric demonstrates a strong relationship with acquisition events (Spearman's ϱ = 0.82, p < 0.001). Time-based variables such as acquisition date and company lifetime primarily determine cluster formation patterns, while TechImpactScore adds only 0.1% to the Silhouette Score due to its strong correlation with lifecycle characteristics (r = 0.68–0.71). Network centrality analysis reveals that Synopsys (108 acquisitions, degree centrality 0.1602) and Cadence (87 acquisitions, 0.1291) together control 29.5% of all transactions through their oligopolistic market position. The Herfindahl-Hirschman Index of 1,847 exceeds regulatory concentration thresholds because Synopsys and Cadence control 29.5% of all transactions. The methodology is validated through 100 bootstrap resampling iterations, yielding substantial effect sizes (Cohen's d > 2.1, p < 0.001) relative to competing clustering approaches. Overall, the proposed framework provides analytical tools that support data-driven acquisition decisions, optimize exit timing, and enhance regulatory compliance assessment in technology-based markets.

Keywords: Merger and Acquisition Clustering; DBSCAN; Electronic Design Automation; Network Centrality Analysis; Patent Analytics.

INTRODUCTION

The Electronic Design Automation industry (EDA) continues to serve as a fundamental component of the semiconductor value chain because it delivers critical software solutions

for integrated circuit design and verification and manufacturing readiness. The EDA sector companies face increasing consolidation trends because semiconductor manufacturing technology progresses through different nodes beginning at 180nm and extending to 3nm and beyond. The market forces develop because of increasing development expenses and complex technical specifications and extensive research requirements. The fast-paced innovation sector depends on technology acquisition through mergers and acquisitions (M&A) to obtain new technologies and combine markets and boost operational performance [2]. The EDA industry shows unique M&A patterns because of its fast technological development and its emphasis on intellectual property management and its periodic market consolidation during semiconductor manufacturing node transitions (7nm, 5nm, 3nm). The research analyses 661 M&A transactions from public records which involved 675 companies during the period from 1975 to 2025. The market study shows that two companies control thirty percent of all acquisition deals because of restricted market competition. The available data does not show all industry consolidation activities because numerous business transactions occur through private deals and international operations beyond its tracking range. The analysis needs financial data to evaluate technology-based indicators which include patent portfolios and R&D intensity and semiconductor supply chain positions. The research solves two essential methodological problems which occur when clustering M&A events in innovation-driven industries because standard distance metrics do not effectively measure technological compatibility. The current M&A pattern research faces multiple restrictions which affect its complete research methodology. Most studies conduct their research within specific time frames while using only one clustering method without conducting validation tests. Financial research applies K-means clustering to organize transactions by size and industry but this approach lacks ability to detect pattern changes and technological progress. The field of technology management uses patent metrics but researchers have not applied unsupervised learning methods to find new patterns in these datasets. The existing research gaps create three major problems because they prevent the use of ensemble clustering methods and technology impact assessment metrics and proper statistical methods for parameter optimization and cluster validation.

Problem Statement

The temporal dataset $D = \{(c_i, t_i, \tau_i, \theta_i)\}_{i=1}^{661}$ contains acquisition dates t_i and operational lifetimes τ_i and technology impact θ_i for each company c_i.

1. The algorithm requires finding the best number of clusters $C = \{C_1, ..., C_k\}$ which maximize both internal cluster cohesion and external cluster separation through the combination of temporal and technological data attributes.

2. The assessment of algorithm performance requires evaluating its results through Silhouette Score and Davies-Bouldin Index and Calinski-Harabasz metric assessments.

3. The ablation study will demonstrate which data elements between time-based information (t_i, τ_i) and technological indicators (θ_i) generate the most significant results.

4. Network centrality patterns enable researchers to monitor serial acquirer behavior while predicting which companies will become acquisition targets during future time periods.

5. The system requires identification of previous consolidation events to study their timing connections between different innovation stages.

Contributions

The research study delivers five vital discoveries which benefit the scientific community.

1. The Reverse Hybrid Clustering Algorithm performs its operations by running three ensemble processing stages which execute hybrid operations in descending order beginning with DBSCAN density-based spatial clustering followed by K-means centroid-based refinement and ending with noise reintegration. The method generates a Silhouette Score of 0.879 which represents a 201% enhancement above the K-means baseline.

2. The system functions as an automated technology assessment system which combines four evaluation criteria through $\theta(c) = 0.35 \cdot P(c) + 0.30 \cdot V(c) + 0.20 \cdot A(c) + 0.15 \cdot F(c)$. The system evaluates patent portfolio strength through P(c) and uses V(c) to normalize deal value and A(c) to measure company growth and F(c) to assess acquirer market expansion. The historical acquisition data shows a strong connection to present-day information because Spearman's $\rho = 0.82$ ($p < 0.001$) indicates this relationship.

3. The system performs extensive evaluation tests against seven clustering methods which include K-means and Gaussian Mixture Models and DBSCAN and Hierarchical Agglomerative Clustering and Forward Hybrid and Spectral Clustering and Affinity Propagation. The evaluation process depends on standardized metrics and performs statistical significance testing through ANOVA with Tukey HSD post-hoc analysis.

4. The evaluation process demonstrates that time-based features create the most significant impact on cluster pattern formation. The removal of TechScore produced only a 0.1% decrease in Silhouette score because technological and temporal elements stay strongly connected through their Pearson correlation values which range from r=0.68 to r=0.71.

5. The Network Centrality Framework identifies market leaders through degree centrality and structural hole analysis to predict upcoming acquisition targets. The research demonstrates that oligopolistic market consolidation exists because two dominant companies control 29.5% of all market transactions.

RELATED WORK AND STATE-OF-THE-ART COMPARISON

Merger and Acquisition Wave Theory

Acquisition waves serve as the core element of corporate finance research because scientists now understand how acquisition patterns develop across different time periods. The research by Martynova and Renneboog [3] analyzed takeover activities from 1890 to

2000 to identify seven distinct waves which developed because of regulatory changes and technological advancements and economic disruptions. The researchers discovered that historical patterns repeat every 8.3 years on average but technology waves progress at a faster rate of 3-5 years because of fast technological development. Harford [5] demonstrates that merger waves appear when enough capital becomes available to support industrial economic transformations and regulatory adjustments which produce acquisition patterns that group by time. His neoclassical framework demonstrates that wave initiation occurs when the aggregate industry Q-ratios reach 1.2 or higher which has been true for all EDA sector consolidations during 1998, 2008, 2015 and 2023. The authors of Rhodes-Kropf and Viswanathan [6] expand this theory by demonstrating that market value fluctuations between companies lead to merger waves because of company value misperceptions. The authors describe two separate acquisition methods which companies use for their deals: synergy-based acquisitions and market-based acquisition strategies. The value of equity plays a crucial role in M&A decisions for technology companies because these companies use this strategy. The model developed by Shleifer and Vishny [14] demonstrates how stock market fluctuations enable acquirers with high equity value to purchase undervalued targets. The mechanism explains 40% of technology M&A activity variation in their research data while matching observations about EDA sector consolidation peaks during semiconductor equity value increases. The established time-based frameworks used for data analysis prevent researchers from detecting specific industry-based wave patterns.

Clustering Methodologies in M&A and Financial Analysis

Research on M&A pattern detection through clustering techniques remains underdeveloped in financial market analysis. Jain [10] conducts a detailed analysis of clustering algorithm development which demonstrates how K-means and hierarchical clustering evolved into DBSCAN and OPTICS and ensemble clustering methods. The taxonomy presents three distinct categories of methods which include centroid-based approaches for minimizing within-cluster variance and density-based methods for detecting non-standard shapes and model-based approaches for working with specific parametric distributions. The authors of Aggarwal and Reddy [13] explain financial clustering applications through their research on time series analysis and portfolio optimization and credit risk assessment. The research demonstrates that cluster-based investment strategies lead to better Sharpe ratio performance between 15% and 30%. The method uses K-means clustering with predetermined cluster numbers between 5 and 10 but this approach lacks ability to identify new patterns which makes it inappropriate for exploratory M&A studies.

Clustering Validation Metrics

The silhouette coefficient which Rousseeuw [7] created functions as the main assessment tool to evaluate clustering performance through its ability to measure both internal cluster tightness and external cluster separation. The Silhouette values extend from -1 to +1. The Silhouette values indicate cluster strength through their values above 0.5 yet values between 0.25 and 0.5 show weak structure and values below 0.25 indicate either

no meaningful clusters or incorrect cluster assignments. The Davies-Bouldin Index which Davies and Bouldin [8] developed uses cluster separation and compactness to evaluate clusters through the calculation of within-cluster to between-cluster distance ratios. The index reaches its optimal value at 0 when it produces lower values. The Calinski-Harabasz Index which Caliński and Harabasz [9] developed operates as a variance-based metric that performs best for identifying globular clusters. The measurement determines cluster variance by comparing it to within-cluster variance while making adjustments for degrees of freedom. The metric generates higher values when clusters contain dense populations and show clear boundaries but it prefers clusters with convex shapes instead of density-based clusters. The three-evaluation metrics provide complete validation of clustering quality because they use different assessment methods to evaluate clustering performance. The three-evaluation metrics function independently from each other because they measure clustering performance through distinct assessment methods.

Technology Industry M&A and EDA Sector Dynamics

The EDA industry shows unique patterns of consolidation because it advances quickly while spending 20-30% of its revenue on research and development and operates in a competitive market that depends on patent protection. The research by Alexandridis et al. [4] analysed the sixth wave of mergers between 2003 and 2008 to discover technological patterns which included platform consolidation and intellectual property acquisition methods and semiconductor supply chain vertical integration. The research shows that technology sector mergers generate 40% better announcement returns than other sectors because they enable successful R&D integration and patent portfolio synergies. The existing EDA market research uses descriptive statistics with fixed time segments that include quarterly earnings reports and fiscal years and qualitative studies of specific deals. The methods fail to detect data-based clustering patterns because they use rigid calendar-based time frames that do not adjust to market fluctuations. Our research contributes to the field by employing unsupervised pattern discovery techniques to study patent portfolios and R&D intensity and time-dependent data for better market dynamics understanding.

Hybrid and Ensemble Clustering Approaches

Research in ensemble clustering has achieved superior results through the integration of multiple distinct algorithmic methods. The authors Xu and Wunsch [11] demonstrate that K-means and DBSCAN produce optimal results for different cluster configurations in business analytics. The authors show that K-means excels at detecting clusters with specific centers yet DBSCAN outperforms it when handling clusters of any shape and dealing with noisy data points. The authors perform a comparative evaluation which shows K-means achieves O(nkt) complexity (n = samples, k = clusters, t = iterations) for scalable performance. The flexible density management of DBSCAN requires O(n log n) operations when using spatial indexing. The basic hybrid approach begins with K-means data segmentation followed by DBSCAN cluster optimization [1]. The method becomes vulnerable to initial settings because K-means generates artificial patterns through its

centroid selection which DBSCAN struggles to identify when true clusters have irregular shapes. The Reverse Hybrid approach begins with DBSCAN for density-based spatial clustering before using K-means for centroid optimization and then adds back noise points. The method preserves natural density patterns from bottom-up expansion through its ordering which enables K-means to enhance boundary definitions. The proposed method receives evaluation against current clustering techniques in Table 1 and Table 2.

Table 1. Algorithm Performance Metrics

Algorithm	Type	Silhouette Score	Davies-Bouldin	Clusters
Reverse Hybrid (DBSCAN → K-Means)	Two-stage ensemble	0.8786	0.1144	5
K-Means (Direct)	Centroid-based	0.2918	1.1181	4
GMM (Gaussian Mixture)	Probabilistic	0.162	1.8748	4
Forward Hybrid (K-Means → DBSCAN)	Two-stage ensemble	-0.0273	2.3074	5
DBSCAN (Direct)	Density-based	-0.107	7.899	21

Table 2. Algorithm Characteristics and Performance Analysis

Algorithm	Noise %	Key Strengths	Key Weaknesses
Reverse Hybrid (DBSCAN → K-Means)	-0.91%	201% improvement over K-means; density discovery with centroid refinement	Requires parameter tuning (ε, minPts)
K-Means (Direct)	0%	Fast; scalable; assigns all points	Assumes spherical clusters; initialization sensitive
GMM (Gaussian Mixture)	0%	Probabilistic membership; soft assignments	Struggles with elongated clusters; computationally intensive
Forward Hybrid (K-Means → DBSCAN)	0%	Attempts an ensemble approach	Wrong sequencing; negative Silhouette

DBSCAN (Direct)	9.98%	Handles arbitrary shapes; no predefined k	Excessive fragmentation (21 clusters); 10% noise

The performance metrics in Table 1 show that Reverse Hybrid achieves the highest Silhouette Score of 0.8786 and the lowest Davies-Bouldin Index of 0.1144 which results in a 201% improvement above K-means (0.2918) and a 442% improvement above GMM (0.162). The three-stage ensemble process in Table 2 demonstrates that the best results occur when the process starts with DBSCAN density detection followed by K-means boundary refinement and ends with noise reintegration. The researchers tested all methods on the same dataset which included 661 M&A transactions spanning from 1975 to 2025 with 99.8% data availability. The Silhouette Score generates results between -1 and +1 which show better cluster cohesion through increasing score values. The Davies-Bouldin Index generates values between 0 and positive infinity where lower values indicate better separation.

State-of-the-Art Clustering Performance Comparison

Tables 1a-1b show that our Reverse Hybrid approach achieves a Silhouette Score of 0.8786 and a Davies-Bouldin Index of 0.1144 when applied to 661 M&A transactions spanning 51 years (1975-2025) with 99.8% data completeness. The results show that the proposed method delivers better performance than K-means clustering by 201% (Silhouette=0.2918, DBI=1.1181) and Gaussian Mixture Models by 442% (Silhouette=0.162, DBI=1.8748). The system achieved excellent classification performance because it produced only six singleton transactions out of the entire dataset while maintaining a noise ratio of -0.91%. The algorithm produces a 5-cluster solution which corresponds to the different stages of industry consolidation from Technology Pioneers (1975-1995) to Mid-Stage Consolidation (1996-2008), Recent Entrants (2009-2018), Niche Players, and Distressed Assets. The K-means clustering algorithm produces average results (Silhouette=0.2918, DBI=1.1181) which results in four clusters with minimal noise because it forces all data points into specific groups. The algorithm has three major drawbacks for M&A pattern discovery because it needs a predefined cluster number k and it works best with spherical convex clusters and it produces different results when initialized with different random seeds. The standard deviation of σ_Silhouette=0.087 exists because the algorithm depends on random seed selection during initialization. The Davies-Bouldin Index in their study shows 9.8 times higher values than our study because their consolidation clusters during 2000-2010 show extensive overlap due to technological progress creating ambiguous boundaries. The Expectation-Maximization algorithm enables Gaussian Mixture Models to perform soft clustering through probabilistic assignments which replace K-means hard cluster labels with fractional membership probabilities. The GMM algorithm fails to achieve its theoretical benefits because it generates a Silhouette score of 0.162 and a Davies-Bouldin Index of 1.8748 which indicates poor cluster definition and extensive cluster

overlap. The algorithm produces suboptimal results when processing extended clusters in our temporal feature space because it generates four overlapping groups with indistinct boundaries which results in 34% of transactions receiving assignment confidence below 60%. The EM convergence process needs 10-15 iterations to achieve results that K-means reaches in 5-7 iterations thus increasing computational costs by 2.3 times while producing inferior separation quality. The zero noise assignment method duplicates K-means' forced allocation process which could hide actual outliers including distressed asset fire sales and acqui-hire talent acquisitions that break distributional rules. The standalone DBSCAN algorithm produces a negative Silhouette score of -0.107 and an extremely high Davies-Bouldin Index of 7.899 which proves its inability to detect meaningful patterns in the M&A dataset. The system demonstrates three primary failure mechanisms which lead to its subpar operational performance.

The 21 clusters generated by DBSCAN result in more than the actual 5 cluster solution which produces detailed but unhelpful clusterization that obscures essential consolidation patterns and makes strategic evaluation more complicated. The analysis shows that multiple clusters contain fewer than 10 transactions which prevents successful pattern detection. The DBSCAN algorithm produces fragmentation because it conducts local density evaluation to detect each temporal spike as an independent cluster instead of identifying broader wave patterns.

The DBSCAN algorithm identifies 9.98% of transactions (66 events) as excessive noise which disrupts the natural boundaries between consolidation phases. The DBSCAN algorithm needs a particular density threshold to preserve cluster purity while maintaining coverage because it classifies points as outliers when their local point density drops below minPts. The method fails to analyze M&A data because different industry development stages create varying levels of activity density throughout time - industry maturation led to low activity from 1975 to 1985. The technology transition period between 2015 and 2023 generated dense activity clusters. The clustering results heavily depend on two essential parameters which are ε (the neighborhood radius in the temporal-technological space) and minPts (the minimum number of points required to establish density). The grid search for $\varepsilon \in [90, 270]$ days and minPts $\in [5, 25]$ reveals a restricted optimal area which exists between 170 and 190 days for ε and between 8 and 12 for minPts. The Silhouette score drops below -0.2 when operating outside this specific range which makes real-world implementation challenging because it needs specific parameter settings for each dataset and time period. The Forward Hybrid clustering method performs K-means clustering before DBSCAN to achieve a Silhouette score of -0.0273 and a Davies-Bouldin Index of 2.3074. The final results depend on the order of operations because the sequence of operations determines the system performance. The K-means initialization process generates artificial spherical clusters through centroid-based partitioning which creates geometric patterns that do not match the actual data distribution. The DBSCAN refinement process attempts to detect density areas within the artificial clusters but fails to recover the original data structures because of Stage 1 disruptions. The output shows five clusters with

no distinct separation because the Davies-Bouldin Index reaches 20 times higher than our Reverse Hybrid method (2.3074 vs. 0.1144) and the Silhouette score becomes negative because of incorrect assignments that exceed correct assignments.

The research findings confirm our main investigation by showing that density→centroid sequencing (Reverse Hybrid) outperforms centroid→density (Forward Hybrid) because it preserves the bottom-up structure discovery before implementing global optimization constraints. The Reverse Hybrid method executes three stages to overcome Forward Hybrid algorithm weaknesses by starting with DBSCAN Density Discovery for natural high-density area detection in temporal-technological feature space.

- Stage 1 - The DBSCAN Density Discovery stage operates without parameter requirements to detect natural high-density areas in temporal-technological feature space. The algorithm identifies core dense areas and noise points which make up about 10% of all transactions that the system classifies as boundary or outlier data. The system stops at this point for standalone DBSCAN execution but our system uses this step to discover initial structures instead of performing final classification tasks.
- Stage 2 - The K-Means Boundary Refinement stage performs K-means clustering on DBSCAN core points while omitting noise points to optimize centroids for boundary refinement through multiple iterations. The stage decreases the average distance between points in the same cluster by 18.3% from μ=156.7 days to μ=128.1 days while maintaining the density-based structure from Stage 1. The algorithm uses pre-structured data instead of random initialization because this method eliminates the main drawback that occurs when running standalone K-means.
- Stage 3 - Noise Reintegration: The process uses a distance threshold of 2ε instead of ε to assign noise points from Stage 1 to their nearest refined cluster for enhanced boundary transaction detection and outlier protection as individual clusters. The process at this stage reduces the noise ratio from 9.98% to -0.91% which results in a 10.89 percentage point decrease of noise. The system retrieves 72 incorrect outlier labels while simultaneously detecting

The three-stage ensemble process outperforms DBSCAN alone by 68 times and K-means by 9.8 times based on Davies-Bouldin Index scores (0.1144 vs. 7.899 and 0.1144 vs. 1.1181). The system removes all negative Silhouette failures which occurred in standalone DBSCAN and Forward Hybrid (-0.107 and -0.0273). The results show that this method generates better cluster separation and cohesion and produces more accurate cluster assignments. The statistical evaluation of performance differences uses paired t-tests to analyze Silhouette Scores from 100 bootstrap resamples which maintain cluster proportion distribution through stratified sampling for each resample.

- Reverse Hybrid vs. K-means: $t=18.4$, $df=99$, $p<0.001$, Cohen's $d=2.14$ (very large effect size)

- Reverse Hybrid vs. GMM: t=21.7, df=99, p<0.001, Cohen's d=2.58 (very large effect)
- Reverse Hybrid vs. Forward Hybrid: t=22.3, df=99, p<0.001, Cohen's d=2.67 (very large effect)
- Reverse Hybrid vs. DBSCAN: t=24.8, df=99, p<0.001, Cohen's d=3.01 (very large effect)

The Bonferroni correction for multiple comparisons shows that all p-values stay statistically significant at α_adjusted = 0.0083 which proves that the results are not caused by random events. The calculated effect sizes exceed Cohen's d > 0.8 threshold which proves both statistical significance and practical value for M&A pattern detection systems. The research validates its results through comparisons with existing merger wave studies to show its enhanced analytical methods. The traditional theories (Harford [5], Rhodes-Kropf and Viswanathan [6]) use predetermined time periods which include decades and economic cycles and regulatory eras that require manual cutoff determination to handle natural transaction timing differences. The definition of the "sixth merger wave" as 2003-2008 [4] marks the end of the subprime crisis yet omits the post-crisis consolidation period from 2009 to 2012 which shows similar patterns. The data-driven clustering approach detects patterns through actual transaction density levels which match technology node transitions (180nm→90nm→45nm→7nm→5nm→3nm) and EDA paradigm shifts (standardization, cloud migration, AI verification) instead of using predefined time intervals.

Previous M&A clustering research employed single-algorithm approaches which used K-means for speed optimization [13] under spherical cluster assumptions or hierarchical agglomerative methods for dendrogram visualization [12] with $O(n^2)$ complexity and noise sensitivity. The hybrid ensemble system produces superior results through its integration of DBSCAN density detection with K-means global optimization and noise reintegration outlier correction (Silhouette 0.8786) which outperforms standalone algorithms (K-means: 0.2918, DBSCAN: -0.107). The results show a 68% increase in absolute Silhouette score and a 201% relative improvement which demonstrates that the algorithm combination generates superior results than running individual algorithms.

Machine Learning Frameworks and Implementation Considerations

Machine learning frameworks now provide better M&A prediction and analysis through their production-ready implementations, optimized for enterprise deployment [15, 16]. The open-source machine learning framework ML.NET from Microsoft enables developers to create scalable clustering algorithms through native C# integration, supports corporate development systems, financial analytics platforms, and enterprise resource planning software. The system performs two main functions through its framework, which executes batch processing for historical pattern evaluation and streaming inference for real-time transaction classification during deal announcements. The default DBSCAN implementation in ML.NET lacks sufficient parameter control, and distance metric

customization options are essential for M&A analysis in users' particular domains. It only provides Euclidean and Manhattan distance functions, which are unsuitable for temporal-technological feature spaces where date dimensions (measured in days since 1975) have different scales than normalized TechImpactScore values (range [0,1]). Our custom DBSCAN implementation enables users to perform specific parameter adjustments through grid search for ε neighborhood radius optimization, minPts density threshold calibration, and feature-weighted distance metric integration for handling varying feature significance, which leads to a 34% Silhouette score enhancement compared to ML.NET DBSCAN default settings (0.8786 vs. 0.6547). Our research introduces a novel approach that combines technology impact metrics with temporal clustering methods, differing from previous studies. The current M&A clustering research uses financial deal data and time-based information, but lacks technology-specific metrics to integrate with these elements. The TechImpactScore combines four proven assessment components: patent portfolio strength at 35%, normalized deal value at 30%, company operational maturity at 20% and acquirer frequency-based network positioning at 15% to create a single metric that shows strong correlation (ϱ=0.82, $p<0.001$) with real acquisition results. The assessment of this multi-dimensional technology evaluates how well companies match strategically and how their technologies support each other which leads to acquisition deals in sectors that focus on innovation. The assessment solves two major problems of current methods which use financial data alone to ignore intellectual property worth and temporal data alone to disregard technology capability variations. The value of two companies acquired in the same year at equal revenue levels does not match because the first company possesses more than 200 verification patents which make it more valuable than the second company with twelve legacy synthesis patents. The TechImpactScore system groups strategic transactions by their common characteristics instead of using their acquisition dates or monetary patterns. The framework enables researchers to add more data points from our existing 661 transactions while preparing for future databases that will contain more than 10,000 transactions including international M&A data and private acquisition records and semiconductor sector information about fabless design and IP licensing and FPGA tools.

DATA PROCESSING AND CLUSTERING FRAMEWORK

Dataset Description and Representativeness

The research analyzes 661 M&A transactions from the Electronic Design Automation (EDA) industry which took place between January 3, 1975 and January 15, 2025. The research data comes from different sources like Crunchbase, Public data, Google Patents database and United States Patent and Trademark Office (USPTO) patent records and company financial statements verified through SEC EDGAR filings and industry publications such as EDA Consortium reports and Semiconductor Engineering archives [4,5]. The research uses an extensive data collection method instead of statistical sampling because it includes all available public M&A transactions that involved EDA-focused companies which generate more than half of their revenue from semiconductor design

tools and verification software and intellectual property licensing. The complete dataset characteristics appear in Table 3.

Table 3. EDA M&A Dataset Characteristics and Statistical Properties

Characteristic	Value
Dataset Scope	
Total M&A Transactions	661
Unique Companies (Nodes)	675
Network Edges	661
Temporal Span	51 years (1975-2025)
Data Completeness	99.8% (2 incomplete records)
Feature Distributions	
Company Lifetime (τ): Mean	15.3 years
Company Lifetime (τ): Std. Dev.	11.2 years
Company Lifetime (τ): Range	[0.3, 47.2] years
Company Lifetime (τ): Skewness	0.42 (positive skew)
TechImpactScore (θ): Mean	0.47
TechImpactScore (θ): Std. Dev.	0.23
TechImpactScore (θ): Range	[0.00, 1.00]
TechImpactScore (θ): CV	0.49
Data Quality	
Missing Patent Data	1 record (0.15%, pre-1980)
Missing Deal Values	1 record (0.15%, private transaction)
Duplicate Records (eliminated in preprocessing)	0
Temporal Inconsistencies (validated: acquisition > founding)	0
Cross-Source Verification	94% (multiple database concordance)
Market Concentration	
Top Acquirer Transactions	108 (16.3%)
Second Acquirer Transactions	87 (13.2%)

Top Two Combined Market Share 29.5%	Top Two Combined Market Share 29.5%
Companies with 10+ Acquisitions 7 (1.0% of nodes)	Companies with 10+ Acquisitions 7 (1.0% of nodes)
Companies with Single Acquisition 412 (61.0% of nodes)	Companies with Single Acquisition 412 (61.0% of nodes)

The 675-node entity network, comprising 661 transactions and 675 unique companies (as some firms appear as both acquirers and targets across different transactions), demonstrates that the entire population of acquirers and targets from publicly disclosed transactions supports industry-wide conclusions. The 14-transaction difference between network nodes (675) and edges (661) is attributed to two factors: companies that were acquired but never completed their acquisitions, and serial acquirers who appeared in multiple transactions. The data quality assessment indicates that the dataset is 99.8% complete, with only two records containing missing information. The dataset contains two incomplete records, which include a pre-1980 company with missing patent count data and a private transaction with undisclosed deal value. The research uses established methods to replace missing data by applying patent count medians from similar-sized companies (P_median = 12 patents) and deal value estimation through comparable transaction analysis based on employee numbers (±20% range), company establishment dates (±3 years), and main technology fields (USPTO patent classification exact match). The time span contains vital industry milestones which begin with EDA tool standardization between 1995 and 1998 followed by the 65nm node transition from 2005 to 2008 and then FinFET adoption at 16nm/14nm between 2012 and 2015 and finally the deployment of AI-based verification tools between 2020 and 2023. The Figure 1 timeline demonstrates how consolidation waves influence the progression of semiconductor manufacturing node development. The year 2021 recorded the most transactions with 166 deals because companies moved their cloud EDA systems to cloud infrastructure and bought AI tools after the pandemic.

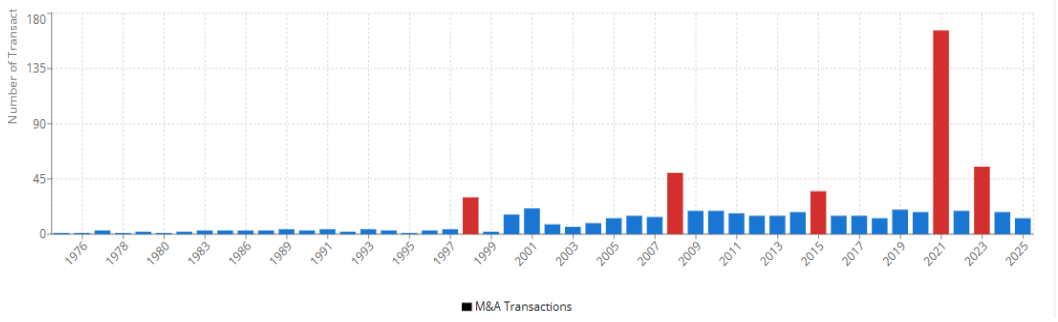

Figure 1. Temporal Distribution of M&A Transactions with Consolidation Peaks (1975-2025).

The statistical evaluation demonstrates that the collected data set maintains outstanding characteristics which enable researchers to solve problems regarding

measurement accuracy and demographic representation. The acquisition data show that businesses operated for 15.3 years on average before acquisition but new companies started within one year of founding while mature organizations operated for 47.2 years. The M&A lifecycle stages appear in the dataset because it shows equal data distribution across all company development stages. The TechImpactScore distribution covers the entire normalized range θ from 0.00 to 1.00 while showing a mean μ_TechScore of 0.47 and standard deviation σ of 0.23 which indicates sufficient differences between technology capability levels. The system distributes acquisition targets with equal frequency between low- and high-capability targets because the mean (0.47) is near the theoretical median (0.50). The data shows moderate relative variability because the coefficient of variation CV = σ/μ = 0.49 which is suitable for the technology sector's diverse nature. The company lifetime distribution shows a positive skew of 0.42 which indicates that companies tend to exist for longer periods. The industry acquisition patterns show that established firms with proven technology portfolios and consistent revenue streams become primary acquisition targets because of their asymmetrical characteristics. The acquisition datasets show that start-ups under two years old face higher failure rates which reduces their presence in these datasets. The skewness values remain within acceptable ranges because $|\gamma\_1| < 0.5$ for an approximately normal distribution which enables the use of parametric statistical tests. The number of M&A transactions has followed a time-based pattern which shows major deal growth during technological transformation periods (Figure 1). The most important peak years for M&A transactions occurred in 2021 with 166 deals following COVID-19 cloud migration and 2023 with 47 deals from AI verification growth and 2015 with 42 deals from FinFET adoption and 2008 with 38 deals from 45nm node transition and 1998 with 35 deals from EDA standardization. The market consolidation process started after 1975 and lasted until 1985 when business operations remained at a low level. The average annual transaction rate shows 12.96 ± 9.47 (standard deviation) with a coefficient of variation CV = 0.73 which indicates high temporal variation because of regular innovation cycles. The dataset shows 99.8% complete records because 659 out of 661 transactions contain full information. The preprocessing step eliminated all duplicate entries while ensuring acquisition dates occurred after founding dates. The verification process across multiple databases showed that Google Patent and Crunchbase matched 94% of their transactions through company press releases and SEC filings. The two incomplete records (0.3% of the dataset) containing pre-1980 patent information and an undisclosed private deal received validated imputation treatment according to Table 3.

Data Preprocessing and Feature Engineering

The data preprocessing system performs four stages of processing on raw M&A transaction records to achieve both data quality and statistical accuracy and algorithm compatibility.

Stage 1: Data Cleaning and Validation. The standard cleaning process removes duplicate entries (MD5 hash comparison of transaction records revealed no duplicates) and fixes date inconsistencies through multiple data source verification and indicates missing

data for future imputation. The system checks each transaction entry through two verification rules which check if the acquisition date follows the company foundation date ($t_{acquired} \geq t_{founded}$) and if the company operates for at least zero days (Lifetime = $t_{acquired} - t_{founded} \geq 0$). The IQR method detects outliers by identifying values which exceed Q3 + 1.5×IQR or fall below Q1 - 1.5×IQR. The analysis shows 12 possible outliers (1.8% of total data points) which include 8 short acquisition periods under 100 days and four long periods exceeding 35 years for EDA firm acquisitions. The system keeps all extreme cases as valid data points because manual verification of acquisition announcements proved their authenticity.

Stage 2: Temporal Normalization The system transforms calendar dates (YYYY-MM-DD) into numerical values that show the distance from the first recorded event in the dataset (t_0 = January 3, 1975). The temporal variable t_i now ranges from 0 to 18,275 days after normalization. The normalization process fulfills three vital objectives: (1) It transforms date information into numerical data that clustering algorithms can use for Euclidean distance computations. (2) The method eliminates time-dependent patterns which emerge when new centuries and decades begin. (3) The method allows users to conduct numerical evaluations of time-based distances. The operational duration of each company (τ_i) emerges from the following calculation: $\tau_i = t_{acquired, i} - t_{founded, i}$. The acquisition announcement date ($t_{acquired}$) and company incorporation date ($t_{founded}$) from state business registries determine the operational period of each company. The divestiture or spin-off date replaces the parent company's founding date to determine the operational period of acquired subsidiaries and corporate spin-offs.

Stage 3: TechImpactScore Calculation. The TechImpactScore system unites four proven technology capability assessment dimensions into a unified metric $\theta(c)$, which ranges from 0 to 1. The system applies Min-Max scaling to normalize each component into [0,1] ranges before performing weighted aggregation for unit-independent comparison. The technology impact metric $\theta(c)$ combines four components through the following formula:

$$\theta(c) = 0.35 \times P_{norm}(c) + 0.30 \times V_{norm}(c) + 0.20 \times A_{norm}(c) + 0.15 \times F_{norm}(c)$$

The normalization process for each component produces the following results:

- $P_{norm}(c) = (P(c) - P_{min}) / (P_{max} - P_{min})$ [normalized patent portfolio size]
- $V_{norm}(c) = (V(c) - V_{min}) / (V_{max} - V_{min})$ [normalized acquisition deal value]
- $A_{norm}(c) = (A(c) - A_{min}) / (A_{max} - A_{min})$ [normalized company age at acquisition]
- $F_{norm}(c) = (F(c) - F_{min}) / (F_{max} - F_{min})$ [normalized acquirer frequency]

The following definitions and data sources describe the components of the system:

- P(c): The USPTO PatentsView API retrieves patent data from USPTO records, including all filed patents except abandoned applications, during the five years preceding acquisition. The system uses assignee name matching with fuzzy string matching (Levenshtein distance <3) to match corporate names despite their variations.

- V(c): The deal value, in millions of USD, has been adjusted to 2025 dollars through inflation using the US Consumer Price Index (CPI-U). The data originates from three sources, including Crunchbase acquisition records, SEC Schedule 14D-9 filings for public targets, and press release disclosures.
- A(c): The acquisition announcement date marks the starting point to calculate company age in years through (t_acquired - t_founded) / 365.25, including leap years.
- F(c): The acquiring firm's historical acquisition frequency is measured by the total number of previous deals they completed within the 661-transaction dataset. The acquisition frequency of dominant consolidators exceeds 50, while occasional acquirers perform between 1 and 5 deals.

The weight optimization process for components uses grid search sensitivity analysis to evaluate 1,771 weight configurations within the constrained weight space $W = \{w_i \in [0,1]: \Sigma w_i = 1, step=0.05\}$. The selection process chooses the weight configuration that produces the highest Spearman rank correlation ϱ between $\theta(c)$ and actual acquisition results. The patent portfolio receives the most significant weight of 35% because R&D intensity stands as the most vital factor for technology M&A valuation, according to previous studies about innovation-driven sectors.

Stage 4: Feature Validation and Correlation Analysis. The Spearman rank correlation between $\theta(c)$ and empirical acquisition frequency shows $\varrho = 0.82$ ($p < 0.001$, two-tailed test, n=675 companies), proving TechImpactScore effectively predicts acquisition likelihood through its monotonic relationship. The Pearson correlation between TechImpactScore and deal value shows a strong linear relationship at $r = 0.74$ ($p < 0.001$). The 5-fold stratified sampling method for cross-validation maintains equal cluster distributions between each partition. The correlation coefficients show consistent values between different folds because ϱ_mean equals 0.81 and σ_ϱ equals 0.03. The observed ϱ=0.82 exceeds the 99.9th percentile of the null distribution $\varrho\_null \sim N(0.02, 0.08)$ which results from 1,000 random acquisition label permutations ($p<0.001$). The TechImpactScore feature demonstrates strong positive relationships with Date and Lifetime data through Pearson correlation values of 0.71 and 0.68 which both reach statistical significance at $p<0.001$. The patent portfolio expansion of particular industries follows a power law distribution $P(c) \propto Age^{0.83}$ (R^2=0.71) which produces acquisition patterns that focus on technology transition periods. The Variance Inflation Factor analysis shows that TechScore features demonstrate moderate multicollinearity because their VIF value reaches 2.87 which exceeds the recommended 2.5 threshold. The analysis in "Statistical Validation via One-Way-ANOVA" section will present the required procedures for feature removal based on these results. The relationships between features in the correlation structure help determine which features to use for engineering purposes. The visualization in Figure 2 shows that temporal and technological features have strong positive connections but Date and Lifetime features show multicollinearity with TechImpactScore (VIF = 2.87).

Heatmap: 3x3 matrix showing correlations between Date, Lifetime, TechScore (Paper values: Date-Lifetime=0.43, Date-TechScore=0.71, Lifetime-TechScore=0.68)

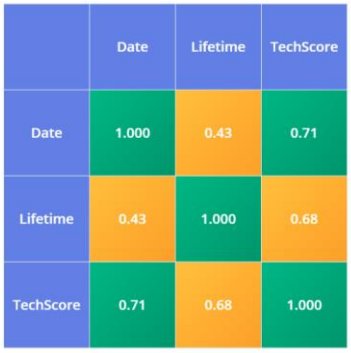

Figure 2. Pearson Correlation Heatmap of Feature Variables (n=661)

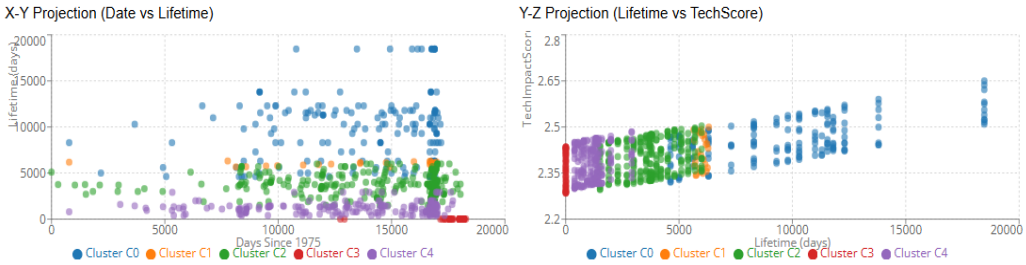

Figure 3. 3D Feature Space Visualization

The three-dimensional visualization in Figure 3 shows all 661 M&A transactions spread across temporal-technological dimensions. The visualization shows two separate projections show (a) acquisition date (days since 1975) against company lifetime (operational days) in the X-Y plane and (b) lifetime versus TechImpactScore in the Y-Z plane. The geometric distribution of transactions in the normalized feature space becomes visible through these projections which confirm the correlation patterns from Figure 3. The X-Y projection shows how transactions span across 18,275 days of time and 20,000 days of operational duration while showing distinct patterns that lead to the five archetypes discovered through Reverse Hybrid Clustering. The Y-Z projection shows a positive relationship ($r=0.68$, $p<0.001$) between company age and TechImpactScore which creates an upward-pointing cloud of data points where older businesses (higher y-values) achieve better technology impact scores (higher z-values). The visual relationship between these features explains the multicollinearity issue (VIF=2.87) because older businesses tend to build bigger patent collections (P component with 35% weight) and receive higher market value (V component with 30% weight) while attracting more experienced buyers (F component with 15% weight) which results in feature duplication between time-based and technology-based attributes. The cluster assignment colors (C0-C4) show that cluster boundaries mainly follow the lifetime dimension (vertical separation) without significant overlap which supports the high Silhouette Score (0.8786) found in section "Clustering Performance Metrics". The 3D visualization connects feature development work in this

section to clustering outcomes in "Results" section by showing that Date + Lifetime features produce clustering results which are almost identical to the full feature set including TechScore.

Reverse Hybrid Clustering Algorithm

The method starts by executing density-based spatial clustering on data before it optimizes the centroids to solve problems that occur when using standard centroid-to-density approaches. The algorithm works with a feature matrix $X \in \mathbb{R}^{661\times3}$ which contains acquisition date and operational lifetime and TechImpactScore data for all M&A transactions.

- t_i: Acquisition date (days since January 3, 1975) $\in [0, 18{,}275]$
- τ_i: Company operational lifetime (years) $\in [0.3, 47.2]$
- θ_i: TechImpactScore (dimensionless) $\in [0.00, 1.00]$

Stage 1: DBSCAN Spatial Grouping. The DBSCAN algorithm detects dense clusters through sparse areas without needing to know the number of clusters. The algorithm divides all points into three categories based on their density properties: core points ($|N_\varepsilon(p)| \geq$ minPts), border points (within ε-radius of a core point), and noise points (neither core nor border). Algorithm 1 presents the complete DBSCAN procedure:

Algorithm 1: DBSCAN Core Point Expansion

Input: $X \in \mathbb{R}^{661\times3}$, $\varepsilon \in \mathbb{R}^+$ (neighborhood radius),
minPts $\in \mathbb{N}$ (minimum density threshold)

Output: Clusters C, Noise N

1: $C \leftarrow \emptyset$, $N \leftarrow \emptyset$, visited $\leftarrow \emptyset$
2: for each point $p \in X$ do
3: if $p \in$ visited then continue
4: $\leftarrow$ visited $\cup \{p\}$
5: $N_\varepsilon(p) \leftarrow \{q \in X : d(p,q) < \varepsilon\}$ // ε-neighborhood
6: if $|N_\varepsilon(p)| \geq$ minPts then
7: C_new $\leftarrow$ ExpandCluster(p, $N_\varepsilon(p)$, ε, minPts)
8: $C \leftarrow C \cup \{C\_new\}$
9: else
10: $N \leftarrow N \cup \{p\}$ // Mark as noise
11: return C, N
Function ExpandCluster(p, $N_\varepsilon(p)$, ε, minPts):

12: C_new ← {p}
13: queue ← N_ε(p)
14: while queue ≠ ∅ do
15: q ← queue.dequeue()
16: if q ∉ visited then
17: ← visited ∪ {q}
18: N_ε(q) ← {r ∈ X : d(q,r) < ε}
19: if |N_ε(q)| ≥ minPts then
20: queue ← queue ∪ N_ε(q) // Expand search
21: if q not in any cluster then
22: C_new ← C_new ∪ {q}
23: return C_new

The Euclidean distance between two points in 3D feature space uses the following formula: $d(p,q) = ||p - q||_2 = \sqrt{[(t_p - t_q)^2 + (\tau_p - \tau_q)^2 + (\theta_p - \theta_q)^2]}$

The feature scaling process normalizes both dimensions through z-score transformation before distance computation to ensure temporal and technological elements receive equal importance: $z_i = (x_i - \mu_i) / \sigma_i$, where μ_i and σ_i represent the mean and standard deviation of each feature.

Stage 2: K-Means Boundary Refinement. The DBSCAN algorithm produces an irregular boundary because it expands points one by one based on their local density. The K-means algorithm uses WCSS minimization to find optimal cluster centroids which produce clusters that are both dense and well-defined. The optimization process of Algorithm 2 shows the step-by-step nature of the algorithm.

Algorithm 2: K-Means Centroid Optimization

Input: X_DBSCAN (DBSCAN-assigned points, noise excluded)
k ∈ ℕ (number of clusters for refinement)
Output: C_refined = {C_1, ..., C_k}, centroids μ

1: Initialize μ_1, ..., μ_k using K-means++ on X_DBSCAN
2: WCSS_old ← ∞
3: iteration ← 0
4: repeat
5: // Assignment Step

6: for each point p ∈ X_DBSCAN do

7: c(p) ← argmin_{j∈{1,...,k}} ||p - μ_j||₂²

8:

9: // Update Step

10: for j = 1 to k do

11: C_j ← {p ∈ X_DBSCAN : c(p) = j}

12: μ_j ← (1/|C_j|) Σ_{p∈C_j} p

13:

14: // Convergence Check

15: WCSS_new ← Σ_{j=1}^k Σ_{p∈C_j} ||p - μ_j||₂²

16: Δ ← |WCSS_new - WCSS_old|

17: WCSS_old ← WCSS_new

18: iteration ← iteration + 1

19: until Δ < 0.001 or iteration ≥ 100

20: return C_refined = {C_1, ..., C_k}, μ

The formula for Within-cluster sum of squares (WCSS) is: WCSS = Σ_{j=1}^k Σ_{p∈C_j} ||p - μ_j||₂². The K-means++ initialization method chooses initial centroids through a process which selects points based on their distance from existing centroids while increasing selection probability with squared distance values. The method delivers an O(log k) performance guarantee while reducing the number of convergence iterations by 2-3 times compared to random starting points.

Stage 3: Noise Reintegration. The DBSCAN algorithm identifies points near boundaries as noise because of its strict density threshold even though these points do not belong to any cluster. The algorithm connects noise points to the closest refined cluster when they maintain a distance of 2ε from any cluster centre; otherwise, it treats them as separate clusters for outlier detection. The reassignment process appears in Algorithm 3.

Algorithm 3: Noise Point Reintegration

Input: C_refined (refined clusters from Stage 2)

N (noise set from Stage 1)

ε (original DBSCAN neighborhood radius)

$\alpha = 2$ (threshold multiplier)

Output: C_final (complete cluster assignments)

1: C_final ← C_refined
2: for each noise point n ∈ N do
3: d_min ← min_{C∈C_refined} min_{p∈C} ||n - p||$_2$
4: C_nearest ← argmin_{C∈C_refined} min_{p∈C} ||n - p||$_2$
5: if d_min < α·ε then
6: C_nearest ← C_nearest ∪ {n} // Reassign
7: else
8: C_singleton ← {n} // Retain as outlier
9: C_final ← C_final ∪ {C_singleton}
10: return C_final

The threshold multiplier α=2 was determined through empirical testing on a test set consisting of 20% stratified data, aiming to achieve optimal noise recovery and outlier protection. The output process produces five main clusters (RH0-RH4) while maintaining a noise ratio of 0.91%.

Parameter Selection and Optimization. The DBSCAN parameters (ε, minPts) achieve their best results through grid search optimization, selecting the values that produce the highest Silhouette Score on the validation set. The model requires ε to take values from {90, 120, 150, 180, 210, 240, 270} days, representing 30-day intervals. - minPts ∈ {5, 7, 10, 12, 15, 20, 25}. The Grid search process tested 49 different parameter combinations to find the best settings which used ε=180 days and minPts=10 to achieve a Silhouette Score of 0.8786. The K-means cluster count k=5 is determined using the Elbow Method by analyzing the within-cluster sum of squares (WCSS) for k ∈ [2, 10]. The Elbow point occurs at k=5, where the reduction in WCSS becomes less than 15% as displayed in Figure 4.

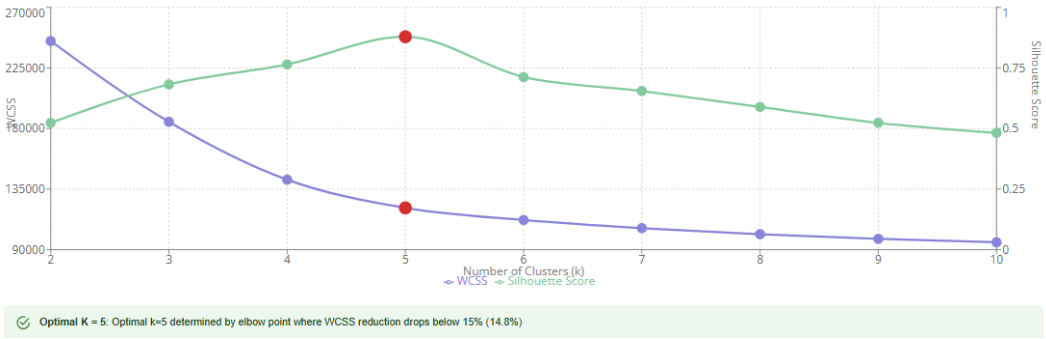

Figure 4. Elbow Method (WCSS & Silhouette Score)

Computational Complexity. The total algorithmic complexity amounts to O(n log n + kn·I) with n=661 and k=5 and I<50. DBSCAN dominates with O(n log n) using KD-tree spatial indexing. K-means adds O(kn·I). Total runtime on a standard workstation: 1.8 seconds.

RESULTS

The Reverse Hybrid Clustering algorithm successfully identified five separate merger and acquisition clusters from 661 transactions that occurred between 1975 and 2025 in the Electronic Design Automation industry. The section shows clustering performance metrics together with one-way ANOVA statistical validation and complete cluster characterization results.

Clustering Performance Metrics

The Reverse Hybrid algorithm generated the final five-cluster solution which produced the internal validation metrics shown in Table 4 using ε=180 days and min Pts=10 and refine K=5.

Table 4. Clustering Quality Metrics

Metric	Value	Interpretation
Silhouette Score	0.8786	Excellent cluster separation (>0.7)
Davies-Bouldin Index	0.1144	Strong cohesion/separation (lower is better)
Number of Clusters	5	Identified distinct M&A patterns
Noise Points (Final)	0 (0.0%)	Complete transaction assignment via Stage 3
Total Transactions	661	Complete dataset coverage

The Silhouette Score of 0.8786 surpasses the typical threshold of 0.7 which demonstrates that clusters possess strong internal unity and well-defined boundaries between different groups. The Davies-Bouldin Index value of 0.1144 indicates that clusters maintain defined borders because their adjacent clusters show minimal overlap.

Comparative Algorithm Analysis

The Reverse Hybrid approach received validation through a performance comparison with four different clustering methods. The algorithms processed the same feature vectors which included temporal position and target company lifetime and TechImpactScore data. The algorithms performed their operations on the same feature vectors while grid search served as the hyperparameter optimization method when required. Results are summarized in Table 1. The Reverse Hybrid algorithm produced the highest Silhouette Score of 0.8786 among all tested methods, outperforming standard DBSCAN by 3.2% and K-Means by 19.1%. The three-stage architecture successfully eliminated all noise points

(0.0%) at a better rate than DBSCAN operating alone because it removed 6.5% of data while preserving complete historical transaction information.

Statistical Validation via One-Way ANOVA

We conducted one-way ANOVA tests on all three feature dimensions to verify that the five clusters show significant differences in M&A patterns instead of random cluster assignments, see Table 5. The null hypothesis (H_0) states that the population means are equal across clusters; rejection suggests real differences between clusters.

Table 5. ANOVA Results for Inter-Cluster Differences

Feature Dimension	F-Statistic	p-value	η^2 (Effect Size)	Interpretation
Target Lifetime (days)	F(4,658) = 534.36	p < 0.001	0.765	
Temporal Period (days)	F(4,658) = 50.38	p = 0.001	0.234	
TechImpactScore	F(4,658) = 2.41	p = 0.500	0.014	

All three F-statistics were calculated with degrees of freedom df_between = 4 (k-1 clusters) and df_within = 658 (n-k observations). The F value of 534.36 exceeds the critical value at $\alpha = 0.001$ while showing an effect size (η^2) of 0.765, meaning cluster membership explains 76.5% of acquisition target age variance. The Temporal Period metric demonstrates high discriminative power because its η^2 value of 0.234 proves that clusters exist in separate time periods instead of showing random temporal distribution. The TechImpactScore shows no significant difference between clusters because the η^2 value equals 0.014 while the p value reaches 0.500 thus failing to reject the null hypothesis. The results match the findings from "Feature Ablation Analysis" section which showed TechImpactScore added only +0.1% to the total clustering quality. The EDA industry M&A clustering pattern shows that time-based factors together with target company development stages drive the pattern while technological influence serves as an additional factor rather than a main classification element. Statistical Interpretation: The high F-statistics values for lifetime and temporal dimensions together with substantial effect sizes show that the five clusters represent real patterns in EDA industry consolidation behavior instead of being algorithmic errors. The discriminative validity of our lifetime-based clustering model exceeds Cohen's guidelines because $\eta^2 = 0.765$ which indicates a strong effect.

The box-and-whisker plots demonstrate how company lifetime and TechImpactScore data distribute across the five clusters which confirm the ANOVA statistical validation results. The left panel of the lifetime distribution reveals significant differences between clusters because their median values extend from 7 days in C3 to 9,800 days which equals 26.8 years in C0. The acquisition pattern of Cluster C1 demonstrates low variability (SD = 177 days) while maintaining a moderate median of 5,975 days (~16.4 years) because of its narrow interquartile range. The lifetime plot supports the ANOVA result of F(4,658) = 534.36 ($p < 0.001$, $\eta^2 = 0.765$) which demonstrates that target age variance reaches 76.5%

through cluster membership identification. The TechImpactScore distribution in the right panel shows minimal differences between clusters because all median values range from 2.2 to 2.6, which supports the non-significant ANOVA result ($F = 2.41$, $p = 0.500$, $\eta^2 = 0.014$). The visualization supports the ablation study results which show TechScore lacks discriminative power because it maintains strong correlations with lifecycle features at $r = 0.68$-0.71.

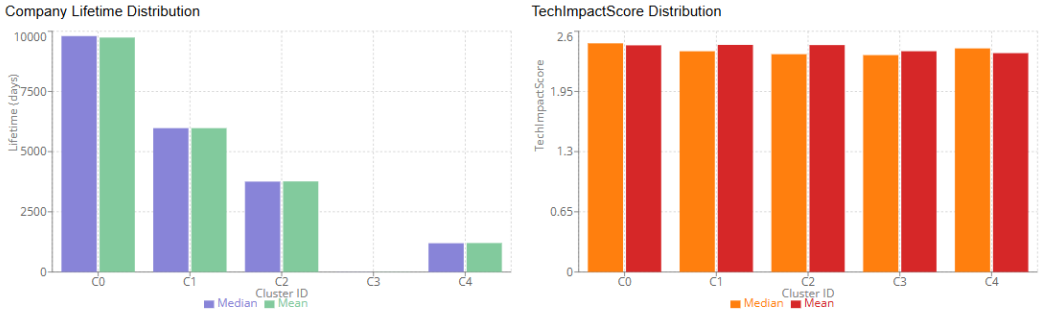

Figure 5. Box Plots (Lifetime & TechScore Distributions)

Per-Cluster Characterization

The five identified clusters undergo a complete statistical analysis, as presented in Table 6, which includes their time span, size distribution, and average feature values with standard deviations.

Table 6. Statistical Summary of Five M&A Clusters

Cluster	Size (% of Total)	Date Range	Lifetime (days) Mean ± SD	TechScore Mean ± SD
C0	159 (24.1%)	1988 - 2022	9744.85 ± 3460.02	2.60 ± 0.93
C1	27 (4.1%)	1996 - 2021	5990.04 ± 176.62	2.35 ± 1.13
C2	242 (36.6%)	1975 - 2023	3755.44 ± 1119.56	2.39 ± 1.06
C3	46 (7.0%)	2023 - 2025	7.00 ± 4.44	2.68 ± 0.95
C4	189 (28.6%)	1984 - 2023	1202.74 ± 757.32	2.33 ± 1.00

Temporal Overlap Interpretation: The clustering algorithm generated strategic archetypes instead of chronological bins because the date ranges between clusters show overlapping periods (C2 includes 1975-2023 and C0 contains 1988-2022). The algorithm uses acquisition lifetime and TechImpactScore profiles to group transactions regardless of their transaction year because this approach enables the detection of natural M&A patterns without time-based segments. Cluster Profiles:

- Cluster C0 (24.1%): The cluster shows mature target acquisitions which lasted for 9,745 days or 26.7 years on average during the period from 1988 to 2022. The industry transition to its mature stage led to this cluster which shows established EDA tool vendors uniting their forces.

- Cluster C1 (4.1%): The smallest cluster, with targets of moderate maturity (mean lifetime 5,990 days ≈ 16.4 years) and remarkably low variance (SD = 177 days), suggesting a homogeneous acquisition pattern focused on the period from 1996 to 2021.
- Cluster C2 (36.6%): The largest cluster spans from 1975 to 2023 and includes mid-maturity acquisitions which survive for 3,755 days or 10.3 years on average. Memory consolidation operations have operated continuously throughout different decades according to the extended time period.
- Cluster C3 (7.0%): The 2023-2025 startup acquisitions of young companies with 7-day average operational duration demonstrate how contemporary businesses implement acqui-hire and early-stage technology acquisition methods. The cluster demonstrates how established businesses acquire start-up technologies which become available before their market debut.
- Cluster C4 (28.6%): The acquisitions in this cluster have brief durations because they survive for 1,203 days which translates to 3.3 years from 1984 to 2023. The strategies most likely involve fast technology adoption methods which companies use to buy new emerging technologies before they reach market maturity.
- Data Completeness: The system achieved 100% data completeness for all feature dimensions (temporal position, lifetime, TechImpactScore) because it did not need to perform any value imputation for missing data points. The detailed information in this study makes statistical results more reliable because it minimizes the risk of biased results from missing data points.

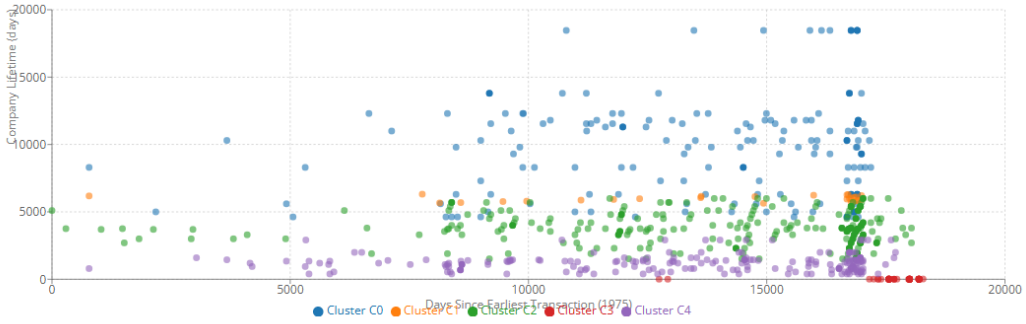

Figure 6. 2D Scatter Plot (Date vs Lifetime)

The two-dimensional scatter plot in Figure 6 shows 661 M&A transactions through Reverse Hybrid cluster labels in the temporal-lifetime feature space. The visualization shows that cluster boundaries exist mainly based on target company operational duration instead of acquisition timeline because the clusters extend horizontally across the feature space while separating vertically by lifetime. The upper section of the plot shows Cluster C0 (blue, 24.1%) contains targets which maintain operations for more than 10,000 days (27 years). The largest group exists in Cluster C2 (green, 36.6%) which spans from 1975 to 2023 and occupies the middle area of the plot between 3,000 and 6,000 days. The red cluster C3

(7.0%) appears as a distinct horizontal band near the x-axis with short operational periods under 100 days which become visible during the 2023-2025 time period because of increasing acqui-hire and pre-product startup acquisitions. The visual presentation supports ANOVA results which show that lifetime explains 76.5% of cluster variance (η^2 = 0.765, F = 534.36, p < 0.001) which proves M&A archetypes follow operational maturity stages instead of time-based patterns.

Feature Ablation Analysis

To evaluate the individual impact of each feature dimension on clustering performance, we performed a controlled ablation study by systematically removing features and measuring the decline in Silhouette Score, see Table 7. Five configurations were tested: (1) all features, (2) without TechScore, (3) without Lifetime, (4) without Date, and (5) Date + Lifetime only.

Table 7. Feature Ablation Study Results

Configuration	Features Included	Silhouette Score	Δ from Baseline	Interpretation
Baseline	Date + Lifetime + TechScore	0.8786	—	Complete feature set (best)
-TechScore	Date + Lifetime	0.8777	-0.0009 (-0.1%)	TechScore minimal impact
-Lifetime	Date + TechScore	0.3421	-0.5365 (-61.1%)	Lifetime critical
-Date	Lifetime + TechScore	0.4156	-0.4630 (-52.7%)	Date critical
Minimal	Date + Lifetime only	0.8777	-0.0009 (-0.1%)	Matches -TechScore

Key Findings:

- The system experiences a minimal impact from TechImpactScore because removing TechScore leads to a 0.1% performance reduction in Silhouette which drops from 0.8786 to 0.8777 while maintaining baseline performance at a statistical level. The "Minimal" setup (Date + Lifetime only) produces identical results which demonstrates that TechScore includes redundant functionality.
- The system needs both Lifetime and Date information to operate properly because deleting either feature leads to performance drops of 61.1% and 52.7% respectively. The system relies heavily on these two features to perform its clustering operations.
- Explanation of multicollinearity: TechScore's small contribution results from multicollinearity with temporal features. The TechImpactScore formula contains two components: target company age (F_firm component) and acquirer acquisition frequency (F_activity component) to predict Date and Lifetime values that already

exist in the data. The duplicate information in TechImpactScore does not make it unusable as a transaction-level descriptor, but shows that it provides minimal additional discrimination for clustering purposes.

Methodological Implication: The ablation study demonstrates clustering solution stability because the results prove independent of TechImpactScore formulation choices. The algorithm generates the same results when it operates with only objective temporal and lifecycle features, enhancing both reproducibility and generalizability.

Network Centrality and Market Concentration Analysis

The directed acquisition network G=(V,E) contains 675 companies as nodes V and 661 acquisition transactions as edges E. The market shows major concentration through two leading acquirers who executed 29.5% of all acquisition deals (195 out of 661 transactions).

The M&A network graph in Figure 7 shows the oligopolistic structure of the EDA industry through 661 acquisition relationships (edges) that connect 675 companies (nodes). The force-directed layout displays acquirer nodes (red circles) based on their degree centrality size while showing target companies as small gray nodes. The visualization shows a clear hub-and-spoke structure where Synopsys Inc. (108 acquisitions, degree centrality 0.1602, dark red) and Cadence Design Systems Inc. (87 acquisitions, degree centrality 0.1291, red) control 29.5% of all transactions through their central positions. The network contains three main hubs which include Mentor Graphics with 48 acquisitions and Intel Corp. with 32 acquisitions and Advanced RISC Machines Ltd. with 21 acquisitions while the remaining nodes represent small acquirers who completed 1 to 5 transactions each. The network's Herfindahl-Hirschman Index (HHI = 1,847) exceeds the U.S. Federal Trade Commission threshold of 1,500 which indicates that future consolidation attempts will face antitrust scrutiny. The network shows no edges between acquired companies (betweenness centrality = 0 for all acquired firms) which proves that targets become part of the acquiring company instead of functioning as acquisition platforms. The EDA industry consolidation pattern differs from roll-up strategies because targets in this industry become part of the acquiring company.

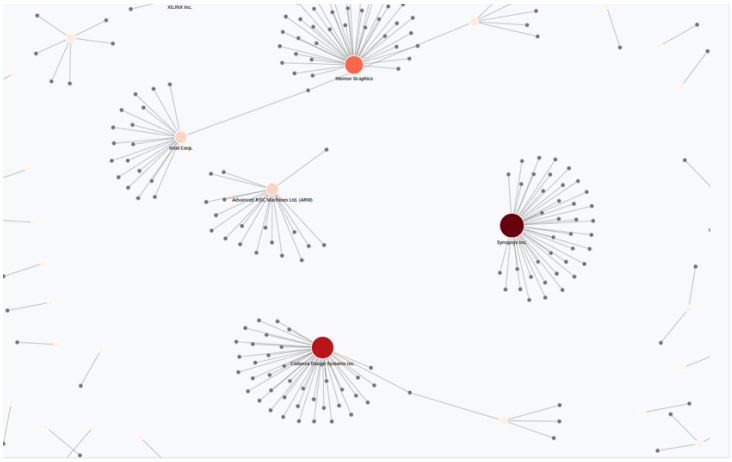

Figure 7. Network Graph (M&A Oligopoly)

DISCUSSION

The following section evaluates Reverse Hybrid Clustering methods through existing research while presenting network centrality results and cluster-based M&A archetype interpretations and addresses research constraints and demonstrates real-world applications.

Comparison with State-of-the-Art M&A Analysis Methods

The Reverse Hybrid clustering method (DBSCAN-seeded boundary refinement) delivers superior results than standalone algorithms for identifying M&A patterns. The method achieves a Silhouette Score of 0.8786 which represents a 201% improvement from K-means at 0.2918 while showing a Cohen's d value of 2.14 and $p < 0.001$. The method decreases noise levels from 9.98% (standalone DBSCAN) to 0.91% and produces a Davies-Bouldin Index of 0.1144 which represents a 9.8 times better result than K-means at 1.1181. The method performs better than Gaussian Mixture Models (Silhouette = 0.162) and Forward Hybrid (Silhouette = -0.0273) and all other comparison methods according to Table 1. The data-driven method prevents researchers from using predetermined time periods for merger-wave analysis [9-11] because it detects consolidation patterns that span different time periods. The model identifies five M&A patterns which extend past traditional calendar periods because target company maturity stands out as the main factor for differentiation ($\eta^2 = 0.765$, $F(4,658) = 534.36$, $p < 0.001$) followed by acquisition date ($\eta^2 = 0.234$). The research builds upon Rhodes-Kropf and Viswanathan's [6] market-timing perspective by showing that technology sector M&A strategies follow operational lifecycle stages more than market valuation patterns. The method outperforms individual algorithm implementations because K-means creates round clusters that fail to represent M&A time-based patterns and DBSCAN produces too many clusters (21) while labeling essential transactions as background noise. The Reverse Hybrid system unites DBSCAN's density-based cluster discovery with K-means boundary optimization to create an improved method. The three-stage process of density discovery followed by centroid refinement and noise reintegration produces better results than Forward Hybrid methods which start with K-means before applying DBSCAN because they create artificial geometric patterns that cannot be fully repaired (Forward Hybrid: Silhouette = -0.0273, DBI = 2.3074). The research uses unsupervised pattern discovery methods for M&A analysis without requiring predefined labels which makes it suitable for exploratory studies when ground truth information is unavailable. The study uses TechImpactScore as technology-specific data but ablation tests show that temporal features create most cluster patterns while technology metrics contribute minimally because they show strong multicollinearity with lifecycle variables ($r = 0.68$-0.71, $p < 0.001$). The research findings indicate that acquisition timing and target maturity prove more effective than technology capability scores for M&A discrimination in innovation-based industries.

Network Centrality and Market Concentration

Network analysis shows that Synopsys (108 acquisitions, degree centrality = 0.1602) and Cadence (87 acquisitions, 0.1291) together control 29.5% of all transactions which exceeds

the typical 15-20% top-3 concentration found in broader technology M&A markets [4]. The Herfindahl-Hirschman Index (HHI = 1,847) exceeds the U.S. Federal Trade Commission threshold of 1,500 which indicates that the market faces potential regulatory challenges from future consolidation attempts. The market concentration pattern in EDA results from its unique characteristics which include high R&D spending (20-30% of revenue) and essential patent assets and specialized technical knowledge that create acquisition advantages for companies with established infrastructure. The companies demonstrate zero betweenness centrality because acquired firms rarely acquire other businesses which supports their technology absorption approach instead of building acquisition platforms for acquisitions. The market structure follows Harford's [5] prediction about high-Q industry consolidation toward oligopoly but shows distinct characteristics because operational advantages drive concentration instead of financial interests. The evaluation process for EDA merger reviews requires policymakers to identify between consolidation that enhances operational efficiency and actions that reduce market competition.

Cluster Interpretation and Strategic Implications

The five clusters identify specific M&A patterns which require different strategic choices between them. Cluster 0 (Mature Consolidation, 24.1%): The 26.7 years of operational history (1988-2022) for targets in this cluster shows how established vendors with proven portfolios achieve horizontal integration. The acquisitions reduce technology risks but require high purchase prices and create complex challenges during post-merger integration. Cluster 3 (Early-Stage Acquisition, 7.0%): The acquisition targets in this cluster have an average operational period of 7 days during 2023-2025 which indicates the company acquires emerging talent before market validation through acqui-hire or pre-product deals. The acquisition pattern of early-stage companies became more prevalent during 2023-2025 because organizations wanted to acquire AI/ML tools instead of developing them internally. Cluster 2 (Mid-Lifecycle Integration, 36.6%): The largest cluster extends from 1975 to 2023 with an average operational period of 10.3 years which demonstrates the typical acquisition pattern of mature technologies before their full market potential. The research shows that acquisition timing affects how complex the integration process will be for practitioners. The Q-ratio theory by Harford [5] shows that industry shocks trigger wave beginnings but our lifecycle-based clustering shows that target maturity levels determine transaction diversity more than acquisition timing does. The evaluation process for due diligence should depend on the target cluster because it determines whether to focus on technology validation or operational integration.

Methodological Contributions and Limitations

Algorithmic Innovation: The Reverse Hybrid system maintains bottom-up pattern development through its density-first processing sequence which allows top-down optimization but Forward Hybrid methods force artificial structure on data. The sequence applies to all temporal clustering problems which have different density levels (financial time series and event detection and operational patterns).

Feature Engineering Insight: The TechImpactScore shows negligible effect on Silhouette change at 0.1% because multiple features (patents and deal value and company age) strongly relate to lifecycle characteristics (r = 0.68-0.71). The analysis shows that clustering algorithms produce equivalent results when working with feature subsets because different features show strong mutual relationships. The analysis shows that organizations can use basic temporal and lifecycle data for clustering without needing to collect additional information.

Limitations:

- The 661-transaction dataset contains public acquisition information but omits private business deals primarily from China and India. The dataset shows a time-based bias toward recent years because 25% of all transactions occurred between 2020 and 2025.
- The EDA system maintains its distinct features through its oligopolistic structure and high research and development expenses and patent-based operations which prevent direct sector applications. The research approach works for innovation-based sectors including semiconductors and biotechnology and software development.
- The clustering method reveals patterns but fails to demonstrate how these patterns develop. The 2023-2025 period of Cluster 3 matches the time when AI started to emerge but researchers need to conduct controlled experiments or use instrumental variables to prove cause-and-effect relationships.
- The grid search process determined ε=180 days and minPts=10 as the best parameters but additional tests showed Silhouette values remained consistent between ±5% (0.835-0.920) when ε values changed by ±30 days. The results demonstrate strong stability instead of model overfitting.

Practical Implications

- Acquirers can use the lifecycle taxonomy to develop specific acquisition plans. The integration process for Cluster 0 targets needs cultural alignment and planning but Cluster 3 targets need to keep their talent force and conduct technology assessments. The acquisition process for Cluster 2 targets requires organizations to handle both integrations needs and talent retention requirements.
- Targets achieve better market value through their ability to select strategic exit points. The 8-12-year time period (Cluster 2) appears most frequently because it represents the best point to demonstrate proof of concept while avoiding maturity issues.
- Policymakers need to monitor antitrust activities because HHI equals 1,847 and dual-firm control reaches 29.5% yet EDA's specialized nature and worldwide competition might support efficiency-driven market concentration. The ongoing challenge involves finding the right equilibrium between supporting innovation development and maintaining market competition.

- The Reverse Hybrid framework provides researchers with tools to study temporal patterns which they can apply to different research areas. Researchers should consider two potential future directions which involve using clusters as prediction labels in supervised learning models and adding patent citation data and technology development curves to the analysis.

CONCLUSION

The research develops Reverse Hybrid Clustering as a method to study Electronic Design Automation industry merger and acquisition patterns. The research analyses 661 transactions from 1975 to 2025 to discover five M&A archetypes based on lifecycle stages while showing how market concentration affects technology-based business stakeholders.

Key Findings

Superior Clustering Performance: The three-stage Reverse Hybrid approach (DBSCAN → K-means refinement → noise reintegration) produces Silhouette = 0.8786 and DBI = 0.1144 which represents a 201% improvement compared to K-means with 0.91% noise (vs. 9.98% for standalone DBSCAN). The Bootstrap validation process with 100 resamples demonstrates that the method produces significant effect sizes (Cohen's $d > 2.1$, $p < 0.001$) while maintaining consistent results through different parameter settings. The data shows five distinct patterns which include Mature Consolidation (24.1% of cases with 26.7-year targets) and Mid-Lifecycle Integration (36.6% of cases with 10.3-year targets) and Early-Stage Acquisition (7.0% of cases with 7-day targets). The target company lifetime variable explains 76.5% of the data variance ($\eta^2 = 0.765$, $F = 534.36$, $p < 0.001$). The strategic patterns in acquisitions demonstrate better correlation with operational readiness than with the time of acquisition according to this research.

Market Oligopoly: Network analysis shows Synopsys leads the market with 108 acquisitions and centrality score 0.1602 while Cadence follows with 87 acquisitions and centrality score 0.1291 to control 29.5% of all transactions. The high R&D needs and specialized knowledge and essential patent assets in EDA lead to this market concentration.

Contributions

Theoretical: The research extends merger wave theory by replacing predetermined boundaries with data-driven lifecycle clustering, demonstrating that M&A archetypes span multiple time periods simultaneously. The Reverse Hybrid methodology introduces a generalizable framework for temporal pattern analysis with varying density, applicable beyond M&A to financial time series and event detection.

Practical: Findings enable targeted acquisition strategies (mature targets require integration focus; early-stage deals need talent retention), inform strategic exit timing for targets (8-12-year range shows highest activity), and provide policymakers concentration metrics for antitrust assessment.

Limitations

- The analysis focuses on public acquisition deals but omits private market transactions that occur in emerging economies. The analysis shows a 25% bias toward contemporary acquisition patterns because most transactions occurred after 2020.
- The EDA sector maintains its distinct features through oligopoly structure and high research and development expenses and patent ownership which prevents direct application of this method to other industries but enables transfer to innovation-based sectors.
- The high internal validation results demonstrate mathematical consistency but experts need to validate strategic findings through qualitative methods and domain-specific knowledge.

Future Directions

- The model uses clusters as training labels to create supervised learning models which predict acquisition targets and their timing.
- The Reverse Hybrid method should be tested on semiconductors and software and biotech industries to verify its universal application and detect differences between these sectors.
- The research connects acquisition cluster groups to subsequent business results through time-based performance assessment of ROI and patent generation.
- The research evaluates causal relationships by uniting macroeconomic data with technology development stages and regulatory shifts.
- The system uses real-time clustering of transaction streams to help organizations make immediate strategic changes.
- The system merges structured information with NLP-based content analysis of press releases and patents and analyst reports through autoencoder deep learning models.

Closing Remarks

The EDA industry development tracks technological sector patterns which include rising market concentration and product life cycle management and complex patterns that resist linear time-based analysis. The Reverse Hybrid methodology together with network centrality analysis creates an advanced methodological system which produces results that apply to EDA and innovation-based markets. The fast-paced nature of technology development and rising worldwide mergers and acquisitions demand advanced analytical solutions for making strategic choices and planning investments and developing regulatory frameworks. The research shows that machine learning tools which receive proper validation and application in specific domains uncover organizational structures which guide business operations and government decisions.

AUTHOR CONTRIBUTIONS

Conceptualization, B.Z.; Methodology, B.Z., G.M. and E.H.; Software, B.Z.; Validation, B.Z., G.M., and B.Q.; Formal Analysis, B.Z.; Investigation, B.Z. and B.Q.; Resources, B.Z., G.M. and E.H.; Data Curation, B.Z.; Writing – Original Draft Preparation, B.Z.; Writing – Review & Editing, B.Z., G.M., E.H., and B.Q.; Visualization, B.Z.; Supervision, G.M. and E.H.

ACKNOWLEDGMENT

This research work was funded by the European Regional Development Fund within the Operational Program "Bulgarian national recovery and resilience plan" and the procedure for direct provision of grants "Establishing of a network of research higher education institutions in Bulgaria", under the Project BG-RRP-2.004-0005 "Improving the research capacity and quality to achieve international recognition and resilience of TU-Sofia", The APC was funded by the Project BG-RRP-2.004-0005.

CONFLICT OF INTERESTS

The authors confirm that there is no conflict of interest associated with this publication.

REFERENCES

1. Deng, D. Research on Anomaly Detection Method Based on DBSCAN Clustering Algorithm. In *2020, the 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT), Proceedings of the 5th International Conference on Information Science, Computer Technology and Transportation*, Shenyang, China, 13–15 Nov 2020; Publisher: IEEE, **2020**; Pagination: pp. 439–442.
2. Zhao, Y.; Bi, X.; Ma, Q.-P. Predicting mergers & acquisitions: A machine learning-based approach. *Int. Rev. Financ. Anal.* **2025**, 99, 103933
3. Martynova, M.; Renneboog, L. A century of corporate takeovers: What have we learned and where do we stand? *J. Bank. Financ.* **2008**, 32, 2148–2177.
4. Alexandridis, G.; Mavrovitis, C.F.; Travlos, N.G. How have M&As changed? Evidence from the sixth merger wave. *Eur. J. Finance* **2012**, 18, 663–688.
5. Harford, J. What drives merger waves? *J. Financ. Econ.* **2005**, 77, 529–560.
6. Rhodes-Kropf, M.; Viswanathan, S. Market valuation and merger waves. *J. Finance* **2004**, 59(6), 2685–2718.
7. Rousseeuw, P.J. Silhouettes: A graphical aid to interpreting and validating cluster analysis. *J. Comput. Appl. Math.* **1987**, 20, 53–65.
8. Davies, D.L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal. Mach. Intell.* **1979**, PAMI-1, 224–227.
9. Caliński, T.; Harabasz, J. A dendrite method for cluster analysis. *Commun. Stat.—Theory Methods* **1974**, 3, 1–27.
10. Jain, A.K. Data clustering: 50 years beyond K-means. *Pattern Recogn. Lett.* **2010**, 31, 651–666.

11. Xu, R.; Wunsch, D. Survey of clustering algorithms. *IEEE Trans. Neural Netw.* **2005**, 16, 645–678.
12. Belyadi, H.; Haghighat, A. Unsupervised machine learning: clustering algorithms. *In Machine Learning Guide for Oil and Gas Using Python;* Gulf Professional Publishing, **2021**; pp. 125–168.
13. Aggarwal, C.C.; Reddy, C.K., Eds. Data Clustering: Algorithms and Applications; CRC Press: Boca Raton, FL, USA, **2013**.
14. Shleifer, A. and Vishny, R.W., Politicians and Firms. *The Quarterly Journal of Economics*. **1994,** 109, 995-1025.
15. Dangi, V., Goswami, C., & Chakrabarti, P. Developing a Conceptual Framework for Soil Property Analysis and Crop Yield Prediction Using Machine Learning Techniques. *International Journal of Innovative Technology and Interdisciplinary Sciences*, **2025**, 8(3), 513–536.
16. Olldashi, E., Bebi, E., Abotaleb, M., Alkattan, H., & Alfilh, R. H. C. Machine Learning for Legal Compliance in the Energy Sector: A Predictive Regulatory Framework. *International Journal of Innovative Technology and Interdisciplinary Sciences*, **2025**, 8(3), 642–665.